4ZA-06

多次元データの可視化を用いた大容量通信からのパケット抽出方法の提案

辻本　真喜子†,　中村　康弘†
防衛大学校 理工学研究科 サイバーセキュリティ工学

1　はじめに

近年，インターネット経由のサイバー攻撃は社会問題になっている [1]．サイバー攻撃への対処は，シグネチャベースのパターンマッチングを用いたリアルタイム分析が有効であることが知られている．一方で，リアルタイム分析による検知を回避するために長期間に渡って行われるサイバー攻撃事例もある．深刻化する攻撃に対処するためには，通常の通信と見分けがつきにくいとされる攻撃の初期段階 [2] を捕捉する必要がある．このため，長期間蓄積した通信の分析（以下，長期間分析）について検討する．

本論文では，シグネチャが存在しない状況下でサイバー攻撃の初期段階を捕捉するための課題を整理した上で，収集したデータから価値のある情報を生成するインテリジェンスサイクル（以下，IC）の考え方を長期間分析に適用したフレームワークを提案する．このフレームワークを 86,400 秒間に渡りセンサが観測した約 28GB のデータセットに適用した結果，サイバー攻撃の初期段階と思われる特徴的なポートスキャンを発見できたので報告する．

2　関連研究

サイバー攻撃を捕捉するための通信パケットの観測およびその分析に関する様々な研究が行われている．

笠間らは，インターネット上で到達可能かつ未使用の IP アドレス空間（以下，ダークネット）へ到達する通信を観測している [3, 4]．ダークネットに到達する通信は攻撃通信の可能性が高いとされ，攻撃者の傾向を把握できるとされる．

既存の侵入検知装置（IDS）は，過去の攻撃通信の特徴をシグネチャとしてリアルタイム分析を行う．リアルタイム分析の結果を可視化し，サイバー攻撃の規模や傾向を可視化する研究もある [5].

3　長期間分析の課題

攻撃者の意図や攻撃通信の存在そのものが不明な状況で，サイバー攻撃の初期活動と判断できる通信を抽出することは困難である．このため，すべての通信を観測した結果から，個々の通信の相互関係の解釈や意味付けを行わなければならない．とくに，明確な意図を持った攻撃者は攻撃の初期活動や走査活動が検知・排除されないように，攻撃通信の低速化や分散化などの様々な工夫を行って一般の通信に紛れ込ませているため，一般によく用いられるようなデータ量の極大値や統計量などの情報だけでは不十分である．

したがって，長期間分析から有益な情報を得るためには，(1) すべての通信を長期間に渡って蓄積・分析しなければならない，(2) 既知ではないパターンを抽出できなければならない，という課題がある．

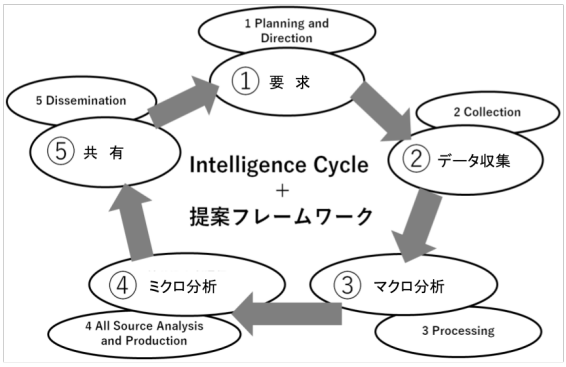

図 1: インテリジェンスサイクルと提案フレームワーク

4　長期間分析のフレームワーク

長期間分析は，既知のシグネチャをリアルタイムに検知する手法と異なり，観測された全通信の中から目的に応じて何らかの有益な情報を抽出できなければならない。このため，詳細に分析するべきパケットを事前にフィルタリングして絞り込むことができない。

このような状況における分析手法のひとつに Intelligence Cycle(以下，IC) がある [6]．IC は，意思決定に必要な情報の要求から提出までの段階をモデル化したものである．*Direction* に基づいて，*Collection* により環境からデータを収集し，情報源の信頼性を加味して *Processing* を行って，*Information* を得る．さらに *Analysis* でそれを解釈・統合して *Intelligence* を生成し，*Dissemination* により，それを活用する．

IC は，収集されたデータから活用可能なインテリジェンスを生成・活用・フィードバックするモデルであり，ここでは，IC を長期間通信観測の分析に適用したフレームワークを提案する（図 1）．サイバー攻撃の初期段

Proposal of an Extraction Method of Characteristic Packets from Large Communication Data Using Multi-dimension Visualization †Makiko Tsujimoto, †Yasuhiro Nakamura, Cyber Security Engineering, National Defense Academy

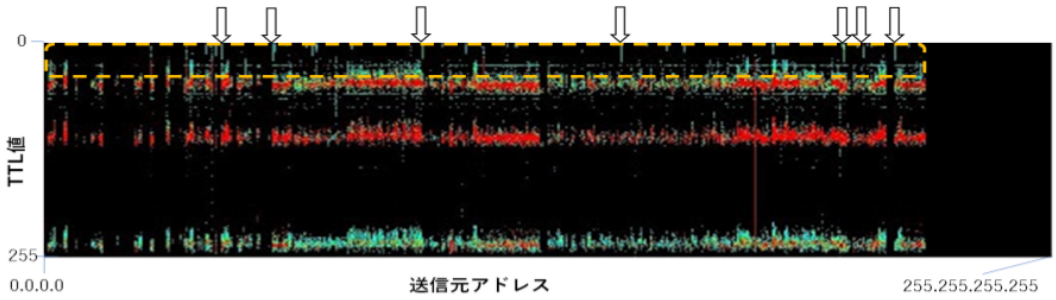

図 2: 可視化結果

階の捕捉を目的（①）としてこのモデルを適用すると，*Collection* は観測可能な全通信の長期間蓄積に相当する（②）．*Processing* は実データを意味のある情報に変換する処理に相当し（③），それらの情報の相互関係や関連性などを調査する作業（④）が Analysis に相当する．また，*Dissemination* は，得られたインテリジェンスをフィルタリングや防御のために活用する活動（⑤）に相当する．以上のことから，IC のモデルは特定のシグネチャを前提とせずに，長期間観測によって得られたデータから走査活動の意味やサイバー攻撃の意図を分析・活用するフレームワークとして有用であると考えられる．

5　適用例

① 要求

大量の通信の中からサイバー攻撃の初期段階あるいは意図的な走査活動と見られる通信を抽出し，その意図を分析する．

② データ収集

防衛大学校のアドレス範囲へ着信した 2013 年以後の全データ [7] を観測データとして用いる．この報告では，そのうち 2019 年 1 月 1 日の観測結果（約 28GB）を処理対象とする．

③ マクロ分析

全データを利用可能な形式に変換する．例えば，各パケットのヘッダ情報フィールドをデータベース化する．ここでは一例として，送信元 IP アドレスと IP ヘッダ中の TTL 値を選択する．

④ ミクロ分析

選択したヘッダ情報フィールドを文献 [8] の方法で可視化する．この結果，特定日時におけるアドレス間の相互関係や特徴量の変化の様子，外れ値などの特徴を確認する．可視化の一例を図 2 に示す．

TTL の初期値は一般に 64,128,255 のいずれかであり，センサまでのホップ数が引かれた値になるはずであるが，複数箇所のアドレス範囲から TTL 値が 1 ~ 12 となる着信があったことがわかる．

このように，攻撃通信の特徴が事前に不明な状態でも，個別フィールドの値を可視化することで特異な通信を視認することができ、さらにペイロードを照合 [7] することで攻撃の意図が推定可能となるものと期待できる．

⑤ 共有

パケットヘッダ特徴などの特異性からアドレス範囲や AS などを調査し，フィルタリングなどの事後処理を行うことでネットワークの運用に活用する．

6　まとめ

本稿では，大容量通信データの長期間分析におけるパケット分析フレームワークを提案し，実データへの適用例を示した．さらに，発信元のアドレス情報や送付されたペイロードを併用することで，攻撃者の意図の推定が可能になるものと期待できる．

参考文献

[1] サイバーセキュリティ対策の推進, 令和元年度版情報通信白書 第 5 節, 2019.
[2] JPCERT, 高度サイバー攻撃への対処におけるログの活用と分析方法 1.1 版, 2016.
[3] 笠間貴弘, NICTER のダークネット長期分析, NICT 研究報告, Vol.62, No.2, 3-1, 2016.
[4] NICT, NICTER 観測レポート 2018, 2018.
[5] NICTERWEB, https://www.nicter.jp/ (2020.1.10).
[6] U.S. Military Joint Chief of Staff, Joint Intelligence, Joint Publication 2-0, I-6, 2013.
[7] 中村康弘, 初期ペイロードに着目したネットワーク走査活動の分析, 情処全大 79, 5D-02, 2017.
[8] 芦野佑樹, 鮫島礼佳, 矢野由紀子, 島成佳, 中村康弘, センサーが捕捉した通信データの解析を支援する可視化手法の提案, SCIS2018, 2018.
[9] 芦野佑樹, 中村康弘, 矢野由紀子, 島成佳, サイバー攻撃の初期段階と推定される活動で使用されるプログラムの分類手法の提案と評価, CSS2017, 2017.